

```

> getwd()
[1] "/Users/binayprasannajena/Documents/GitHub/dsc520"
> dir()
[1] "assignments" "completed" "data" "LICENSE"
"README.md" "RMarkdown.md"
> setwd("/Users/binayprasannajena/Documents/GitHub/dsc520/")
> survey_df <- read.csv(file='data/acs-14-1yr-
s0201.csv',stringsAsFactors = TRUE)
> survey_df[1:8]

```

	Id	Id2	
Geography	PopGroupID	POPGR0UP.display.label	RacesReported
1 0500000US01073	1073		Jefferson County,
Alabama	1	Total population	660793
2 0500000US04013	4013		Maricopa County,
Arizona	1	Total population	4087191
3 0500000US04019	4019		Pima County,
Arizona	1	Total population	1004516
4 0500000US06001	6001		Alameda County,
California	1	Total population	1610921
5 0500000US06013	6013		Contra Costa County,
California	1	Total population	1111339
6 0500000US06019	6019		Fresno County,
California	1	Total population	965974
7 0500000US06029	6029		Kern County,
California	1	Total population	874589
8 0500000US06037	6037		Los Angeles County,
California	1	Total population	10116705
9 0500000US06059	6059		Orange County,
California	1	Total population	3145515
10 0500000US06065	6065		Riverside County,
California	1	Total population	2329271
11 0500000US06067	6067		Sacramento County,
California	1	Total population	1482026
12 0500000US06071	6071		San Bernardino County,
California	1	Total population	2112619
13 0500000US06073	6073		San Diego County,
California	1	Total population	3263431
14 0500000US06075	6075		San Francisco County,
California	1	Total population	852469
15 0500000US06077	6077		San Joaquin County,
California	1	Total population	715597
16 0500000US06081	6081		San Mateo County,
California	1	Total population	758581
17 0500000US06085	6085		Santa Clara County,
California	1	Total population	1894605
18 0500000US06097	6097		Sonoma County,
California	1	Total population	500292

19	05000000US06099	6099	Stanislaus County,
California	1	Total population	531997
20	05000000US06111	6111	Ventura County,
California	1	Total population	846178
21	05000000US08005	8005	Arapahoe County,
Colorado	1	Total population	618821
22	05000000US08031	8031	Denver County,
Colorado	1	Total population	663862
23	05000000US08041	8041	El Paso County,
Colorado	1	Total population	663519
24	05000000US08059	8059	Jefferson County,
Colorado	1	Total population	558503
25	05000000US09001	9001	Fairfield County,
Connecticut	1	Total population	945438
26	05000000US09003	9003	Hartford County,
Connecticut	1	Total population	897985
27	05000000US09009	9009	New Haven County,
Connecticut	1	Total population	861277
28	05000000US10003	10003	New Castle County,
Delaware	1	Total population	552778
29	05000000US11001	11001	District of Columbia, District of
Columbia	1	Total population	658893
30	05000000US12009	12009	Brevard County,
Florida	1	Total population	556885
31	05000000US12011	12011	Broward County,
Florida	1	Total population	1869235
32	05000000US12031	12031	Duval County,
Florida	1	Total population	897698
33	05000000US12057	12057	Hillsborough County,
Florida	1	Total population	1316298
34	05000000US12071	12071	Lee County,
Florida	1	Total population	679513
35	05000000US12086	12086	Miami-Dade County,
Florida	1	Total population	2662874
36	05000000US12095	12095	Orange County,
Florida	1	Total population	1253001
37	05000000US12099	12099	Palm Beach County,
Florida	1	Total population	1397710
38	05000000US12103	12103	Pinellas County,
Florida	1	Total population	938098
39	05000000US12105	12105	Polk County,
Florida	1	Total population	634638
40	05000000US12127	12127	Volusia County,
Florida	1	Total population	507531
41	05000000US13067	13067	Cobb County,
Georgia	1	Total population	730981
42	05000000US13089	13089	DeKalb County,
Georgia	1	Total population	722161

43	05000000US13121	13121		Fulton County,
Georgia	1	Total population		996319
44	05000000US13135	13135		Gwinnett County,
Georgia	1	Total population		877922
45	05000000US15003	15003		Honolulu County,
Hawaii	1	Total population		991788
46	05000000US17031	17031		Cook County,
Illinois	1	Total population		5246456
47	05000000US17043	17043		DuPage County,
Illinois	1	Total population		932708
48	05000000US17089	17089		Kane County,
Illinois	1	Total population		527306
49	05000000US17097	17097		Lake County,
Illinois	1	Total population		705186
50	05000000US17197	17197		Will County,
Illinois	1	Total population		685419
51	05000000US18097	18097		Marion County,
Indiana	1	Total population		934243
52	05000000US20091	20091		Johnson County,
Kansas	1	Total population		574272
53	05000000US20173	20173		Sedgwick County,
Kansas	1	Total population		508803
54	05000000US21111	21111		Jefferson County,
Kentucky	1	Total population		760026
55	05000000US24003	24003		Anne Arundel County,
Maryland	1	Total population		560133
56	05000000US24005	24005		Baltimore County,
Maryland	1	Total population		826925
57	05000000US24031	24031		Montgomery County,
Maryland	1	Total population		1030447
58	05000000US24033	24033		Prince George's County,
Maryland	1	Total population		904430
59	05000000US24510	24510		Baltimore city,
Maryland	1	Total population		622793
60	05000000US25005	25005		Bristol County,
Massachusetts	1	Total population		554194
61	05000000US25009	25009		Essex County,
Massachusetts	1	Total population		769091
62	05000000US25017	25017		Middlesex County,
Massachusetts	1	Total population		1570315
63	05000000US25021	25021		Norfolk County,
Massachusetts	1	Total population		692254
64	05000000US25023	25023		Plymouth County,
Massachusetts	1	Total population		507022
65	05000000US25025	25025		Suffolk County,
Massachusetts	1	Total population		767254
66	05000000US25027	25027		Worcester County,
Massachusetts	1	Total population		813475

67	05000000US26081	26081	Kent County,
Michigan	1	Total population	629237
68	05000000US26099	26099	Macomb County,
Michigan	1	Total population	860112
69	05000000US26125	26125	Oakland County,
Michigan	1	Total population	1237868
70	05000000US26163	26163	Wayne County,
Michigan	1	Total population	1764804
71	05000000US27053	27053	Hennepin County,
Minnesota	1	Total population	1212064
72	05000000US27123	27123	Ramsey County,
Minnesota	1	Total population	532655
73	05000000US29095	29095	Jackson County,
Missouri	1	Total population	683191
74	05000000US29189	29189	St. Louis County,
Missouri	1	Total population	1001876
75	05000000US31055	31055	Douglas County,
Nebraska	1	Total population	543244
76	05000000US32003	32003	Clark County,
Nevada	1	Total population	2069681
77	05000000US34003	34003	Bergen County, New
Jersey	1	Total population	933572
78	05000000US34007	34007	Camden County, New
Jersey	1	Total population	511038
79	05000000US34013	34013	Essex County, New
Jersey	1	Total population	795723
80	05000000US34017	34017	Hudson County, New
Jersey	1	Total population	669115
81	05000000US34023	34023	Middlesex County, New
Jersey	1	Total population	836297
82	05000000US34025	34025	Monmouth County, New
Jersey	1	Total population	629279
83	05000000US34029	34029	Ocean County, New
Jersey	1	Total population	586301
84	05000000US34031	34031	Passaic County, New
Jersey	1	Total population	508856
85	05000000US34039	34039	Union County, New
Jersey	1	Total population	552939
86	05000000US35001	35001	Bernalillo County, New
Mexico	1	Total population	675551
87	05000000US36005	36005	Bronx County, New
York	1	Total population	1438159
88	05000000US36029	36029	Erie County, New
York	1	Total population	922835
89	05000000US36047	36047	Kings County, New
York	1	Total population	2621793
90	05000000US36055	36055	Monroe County, New
York	1	Total population	749857

91	05000000US36059	36059	Nassau County, New
York	1	Total population	1358627
92	05000000US36061	36061	New York County, New
York	1	Total population	1636268
93	05000000US36081	36081	Queens County, New
York	1	Total population	2321580
94	05000000US36103	36103	Suffolk County, New
York	1	Total population	1502968
95	05000000US36119	36119	Westchester County, New
York	1	Total population	972634
96	05000000US37081	37081	Guilford County, North
Carolina	1	Total population	512119
97	05000000US37119	37119	Mecklenburg County, North
Carolina	1	Total population	1012539
98	05000000US37183	37183	Wake County, North
Carolina	1	Total population	998691
99	05000000US39035	39035	Cuyahoga County,
Ohio	1	Total population	1259828
100	05000000US39049	39049	Franklin County,
Ohio	1	Total population	1231393
101	05000000US39061	39061	Hamilton County,
Ohio	1	Total population	806631
102	05000000US39113	39113	Montgomery County,
Ohio	1	Total population	533116
103	05000000US39153	39153	Summit County,
Ohio	1	Total population	541943
104	05000000US40109	40109	Oklahoma County,
Oklahoma	1	Total population	766215
105	05000000US40143	40143	Tulsa County,
Oklahoma	1	Total population	629598
106	05000000US41051	41051	Multnomah County,
Oregon	1	Total population	776712
107	05000000US41067	41067	Washington County,
Oregon	1	Total population	562998
108	05000000US42003	42003	Allegheny County,
Pennsylvania	1	Total population	1231255
109	05000000US42017	42017	Bucks County,
Pennsylvania	1	Total population	626685
110	05000000US42029	42029	Chester County,
Pennsylvania	1	Total population	512784
111	05000000US42045	42045	Delaware County,
Pennsylvania	1	Total population	562960
112	05000000US42071	42071	Lancaster County,
Pennsylvania	1	Total population	533320
113	05000000US42091	42091	Montgomery County,
Pennsylvania	1	Total population	816857
114	05000000US42101	42101	Philadelphia County,
Pennsylvania	1	Total population	1560297

115	05000000US44007	44007	Providence County, Rhode
	Island	1	Total population 631974
116	05000000US47037	47037	Davidson County,
	Tennessee	1	Total population 668347
117	05000000US47157	47157	Shelby County,
	Tennessee	1	Total population 938803
118	05000000US48029	48029	Bexar County,
	Texas	1	Total population 1855866
119	05000000US48085	48085	Collin County,
	Texas	1	Total population 885241
120	05000000US48113	48113	Dallas County,
	Texas	1	Total population 2518638
121	05000000US48121	48121	Denton County,
	Texas	1	Total population 753363
122	05000000US48141	48141	El Paso County,
	Texas	1	Total population 833487
123	05000000US48157	48157	Fort Bend County,
	Texas	1	Total population 685345
124	05000000US48201	48201	Harris County,
	Texas	1	Total population 4441370
125	05000000US48215	48215	Hidalgo County,
	Texas	1	Total population 831073

	HSDegree	BachDegree
1	89.1	30.5
2	86.8	30.2
3	88.0	30.8
4	86.9	42.8
5	88.8	39.7
6	73.6	19.7
7	74.5	15.4
8	77.5	30.3
9	84.6	38.0
10	80.6	20.7
11	86.8	28.9
12	78.6	18.9
13	86.6	37.1
14	88.1	54.2
15	77.6	18.3
16	88.1	47.5
17	87.4	48.4
18	87.6	34.8
19	78.4	17.0
20	83.6	31.6
21	91.9	40.9
22	85.5	44.3
23	92.8	36.5
24	94.1	42.0
25	89.8	46.7

26	89.3	36.8
27	89.5	34.5
28	90.1	35.8
29	90.2	55.0
30	91.6	27.2
31	88.4	30.5
32	89.0	26.1
33	87.3	29.8
34	86.3	26.5
35	80.9	26.6
36	87.9	31.4
37	87.7	33.0
38	90.1	29.5
39	84.9	19.7
40	88.9	22.5
41	90.3	43.7
42	88.4	41.7
43	91.3	49.2
44	88.0	35.4
45	91.8	32.6
46	85.5	36.2
47	92.3	48.0
48	82.9	32.6
49	90.3	44.0
50	90.7	33.1
51	85.0	28.8
52	95.5	52.8
53	88.8	30.7
54	88.5	31.6
55	91.9	38.8
56	90.4	37.2
57	90.9	58.5
58	85.5	31.0
59	84.4	30.0
60	82.5	25.7
61	89.1	38.9
62	92.3	52.3
63	94.1	51.9
64	92.2	34.1
65	83.9	42.3
66	90.1	34.6
67	89.1	33.7
68	89.3	23.9
69	93.6	44.8
70	84.9	22.1
71	93.2	47.3
72	89.9	40.9
73	90.0	29.5

74	93.2	42.8
75	88.2	36.3
76	84.5	22.7
77	91.5	46.2
78	88.3	31.3
79	85.5	32.7
80	83.4	38.2
81	89.1	41.0
82	93.1	43.7
83	91.7	28.6
84	83.8	28.6
85	86.2	33.0
86	88.0	32.7
87	70.5	19.3
88	90.6	31.3
89	80.0	34.3
90	90.3	35.9
91	90.7	43.2
92	86.8	59.9
93	80.4	29.8
94	89.8	34.0
95	87.4	47.1
96	89.0	33.3
97	89.5	43.0
98	92.4	49.2
99	88.1	31.0
100	90.0	38.0
101	90.5	35.6
102	89.7	25.7
103	91.1	30.3
104	86.8	30.6
105	88.6	30.7
106	91.1	41.6
107	90.2	39.7
108	93.9	37.7
109	93.9	37.7
110	92.3	49.3
111	91.5	36.3
112	84.9	26.0
113	93.7	47.3
114	82.6	26.0
115	82.0	25.2
116	86.7	37.3
117	87.4	29.9
118	83.0	26.3
119	93.7	50.0
120	77.6	29.1
121	91.9	41.5


```

122      75.8      21.1
123      88.6      44.1
124      79.8      29.7
125      62.2      17.9
[ reached 'max' / getOption("max.print") -- omitted 11 rows ]
> str(survey_df)
'data.frame': 136 obs. of 8 variables:
 $ Id      : Factor w/ 136 levels
"05000000US01073",...: 1 2 3 4 5 6 7 8 9 10 ...
 $ Id2     : int  1073 4013 4019 6001 6013 6019
6029 6037 6059 6065 ...
 $ Geography : Factor w/ 136 levels "Alameda County,
California",...: 56 70 98 1 20 43 62 68 92 106 ...
 $ PopGroupID : int  1 1 1 1 1 1 1 1 1 1 ...
 $ POPGROUP.display.label: Factor w/ 1 level "Total population":
1 1 1 1 1 1 1 1 1 1 ...
 $ RacesReported : int  660793 4087191 1004516 1610921
1111339 965974 874589 10116705 3145515 2329271 ...
 $ HSDegree : num  89.1 86.8 88 86.9 88.8 73.6 74.5
77.5 84.6 80.6 ...
 $ BachDegree : num  30.5 30.2 30.8 42.8 39.7 19.7
15.4 30.3 38 20.7 ...

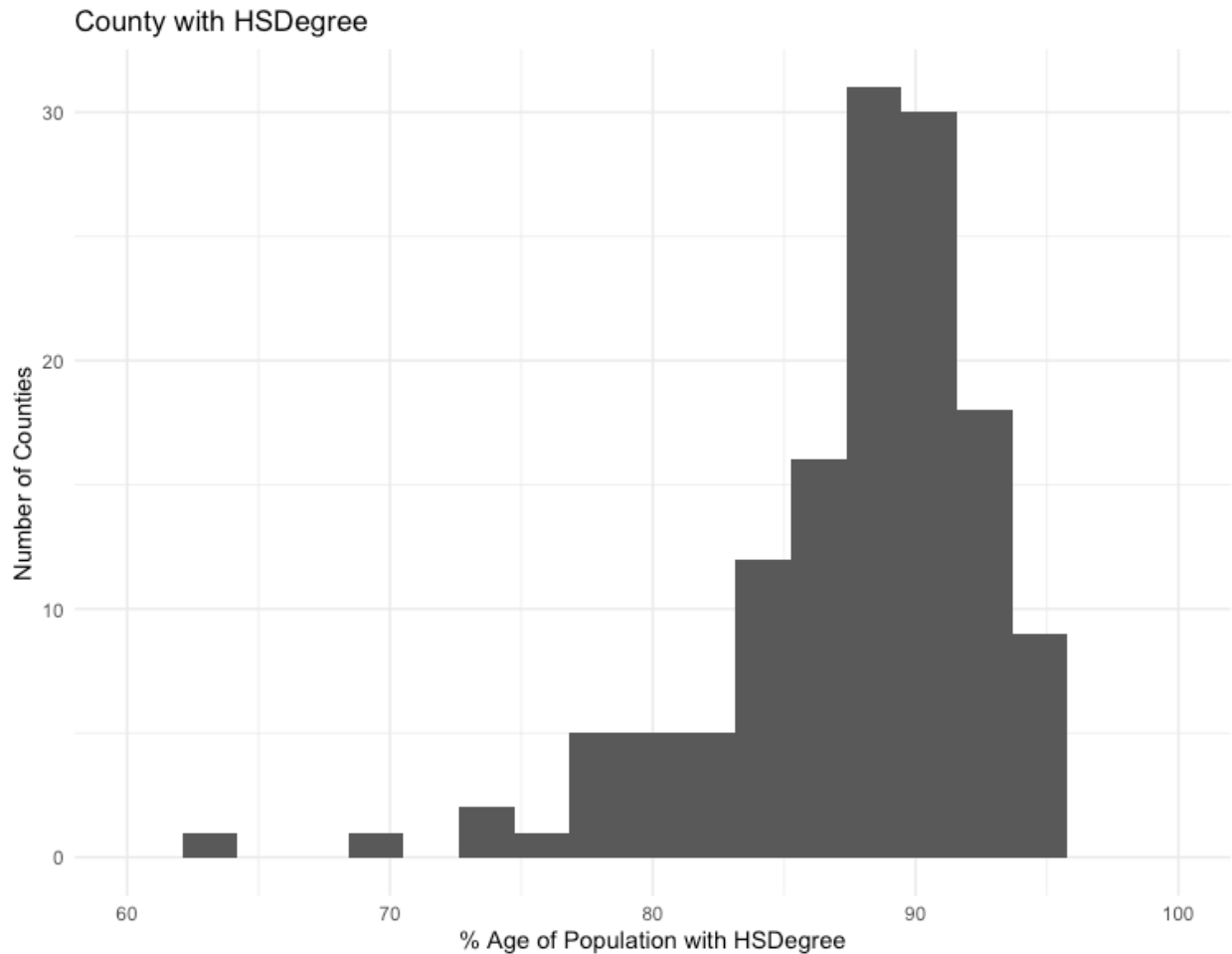
```

```

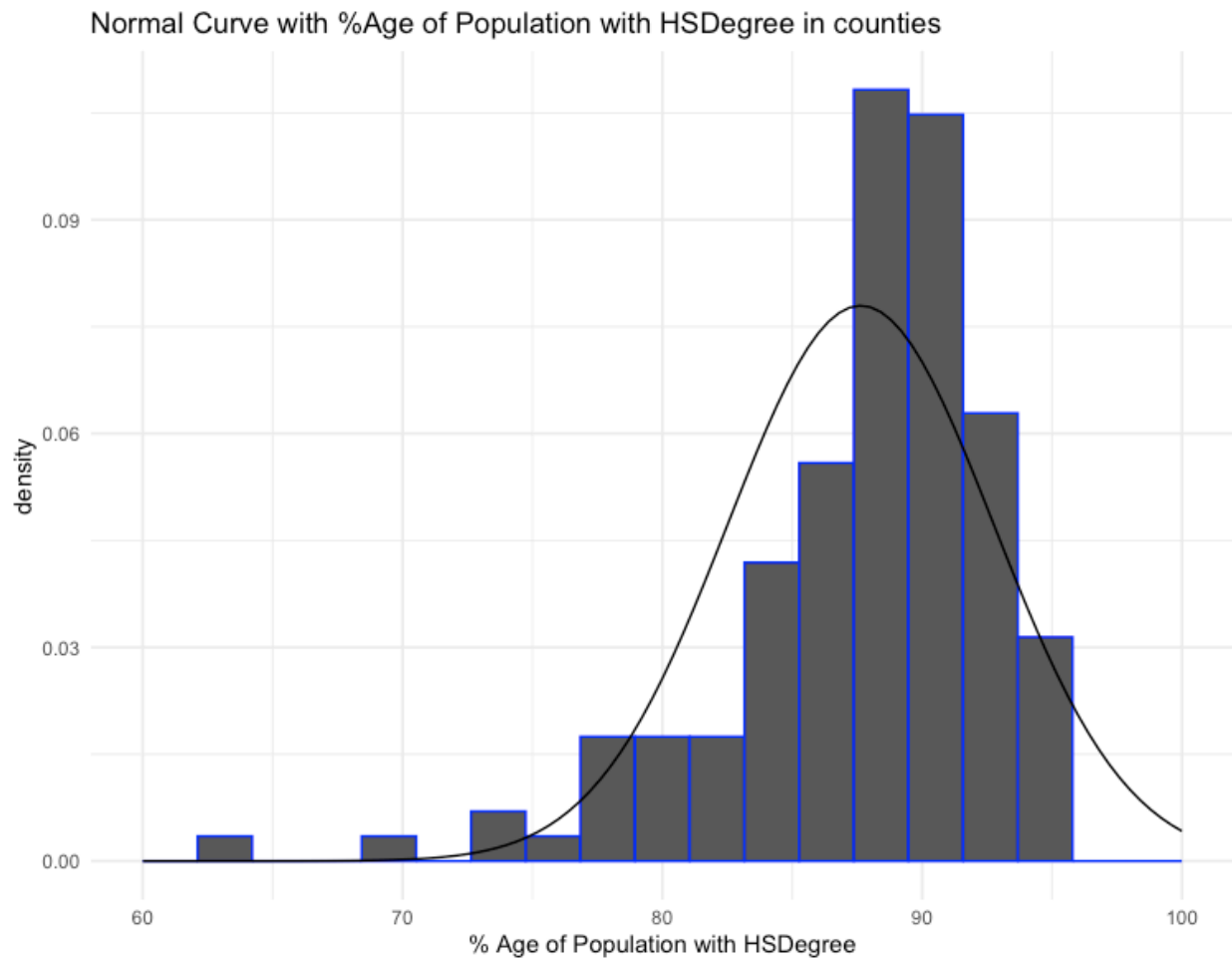
> nrow(survey_df)
[1] 136
> ncol(survey_df)
[1] 8

```

```
> library(ggplot2)
> theme_set(theme_minimal())
> nOfBins <- 20
> ggplot(survey_df, aes(HSDegree)) + xlim(c(60,100)) +
  geom_histogram(bins = nOfBins) + ggtitle('County with HSDegree')
+ xlab('% Age of Population with HSDegree') + ylab('Number of
Counties')
```



```
> ggplot(survey_df, aes(HSDegree)) + xlim(c(60,100)) +  
geom_histogram(aes(y = ..density..), bins=nOfBins, colour =  
"blue") + stat_function(fun = dnorm, args = list(mean =  
mean(survey_df$HSDegree), sd = sd(survey_df$HSDegree))) +  
ggtitle('Normal Curve with %Age of Population with HSDegree in  
counties') + xlab('% Age of Population with HSDegree')
```



```
> summary(survey_df)
```

```
      Id      Id2
Geography PopGroupID POPGROUP.display.label
0500000US01073: 1   Min.    : 1073   Alameda County, California
: 1   Min.    :1   Total population:136
0500000US04013: 1   1st Qu.:12082   Allegheny County,
Pennsylvania: 1   1st Qu.:1
0500000US04019: 1   Median :26112   Anne Arundel County,
Maryland : 1   Median :1
0500000US06001: 1   Mean    :26833   Arapahoe County, Colorado
: 1   Mean    :1
0500000US06013: 1   3rd Qu.:39123   Baltimore city, Maryland
: 1   3rd Qu.:1
0500000US06019: 1   Max.    :55079   Baltimore County, Maryland
: 1   Max.    :1
(Other)      :130                (Other)
```

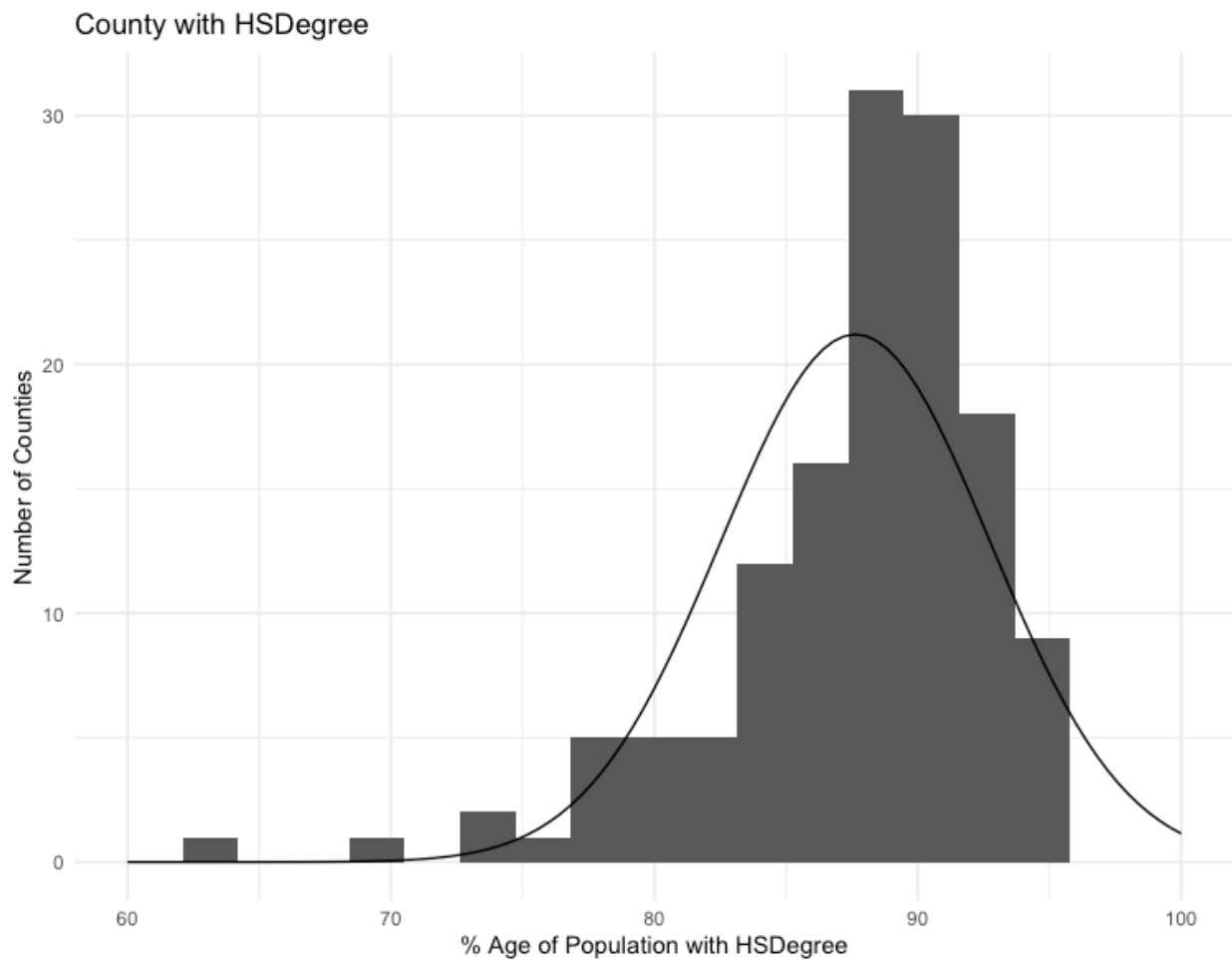
```
:130
```

```
RacesReported      HSDegree      BachDegree
Min.    : 500292   Min.    :62.20   Min.    :15.40
1st Qu.: 631380   1st Qu.:85.50   1st Qu.:29.65
Median : 832708   Median :88.70   Median :34.10
Mean    : 1144401   Mean    :87.63   Mean    :35.46
3rd Qu.: 1216862   3rd Qu.:90.75   3rd Qu.:42.08
Max.    :10116705   Max.    :95.50   Max.    :60.30
```

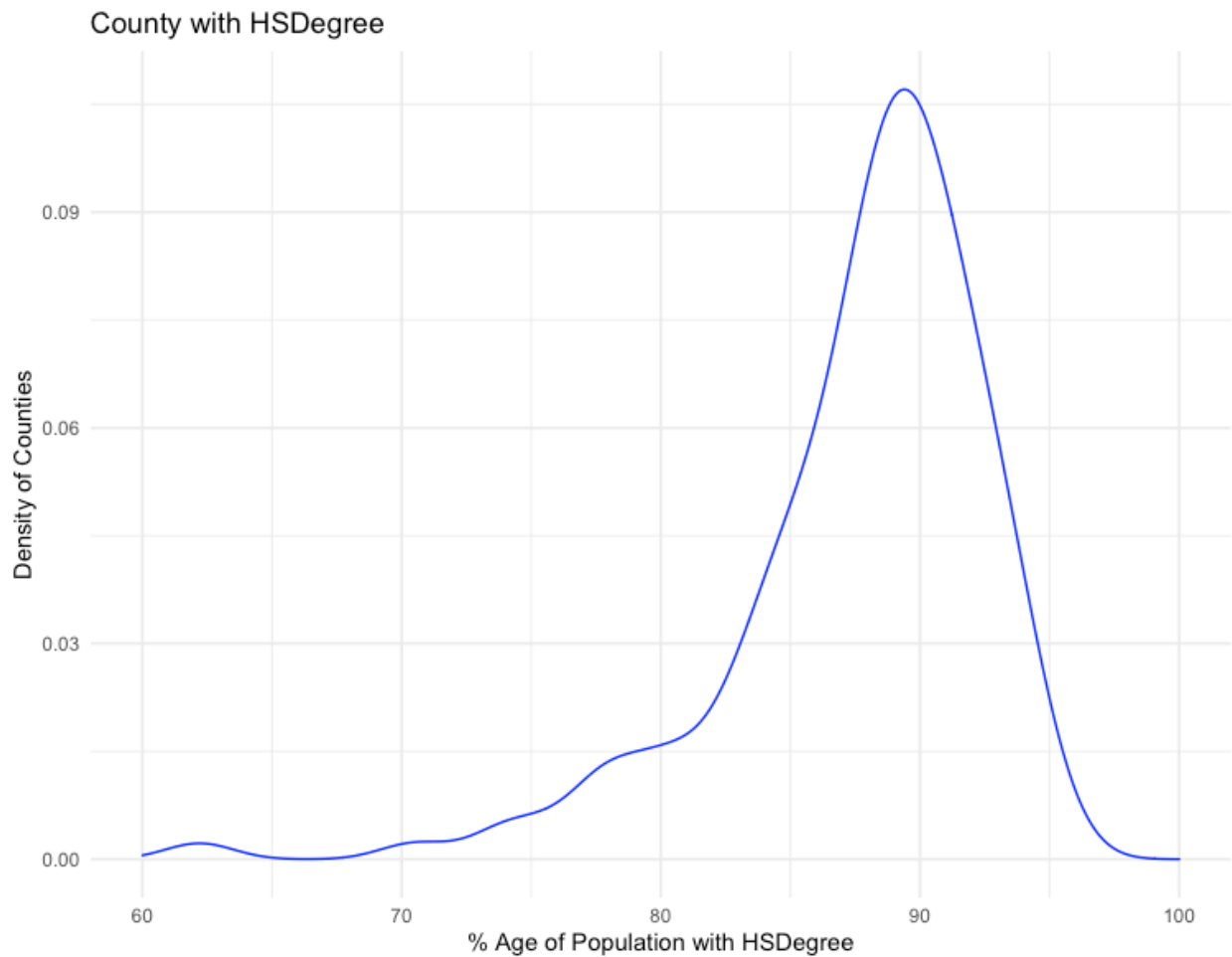
```

> n_obs = sum(!is.na(survey_df$HSDegree))
> ggplot(survey_df, aes(HSDegree)) + xlim(c(60,100)) +
  geom_histogram(bins = nOfBins)+ ggtitle('County with HSDegree')
+ xlab('% Age of Population with HSDegree') + ylab('Number of
Counties') + stat_function(fun = function(x) dnorm(x, mean =
mean(survey_df$HSDegree), sd = sd(survey_df$HSDegree)) *
(40/nOfBins) * n_obs)

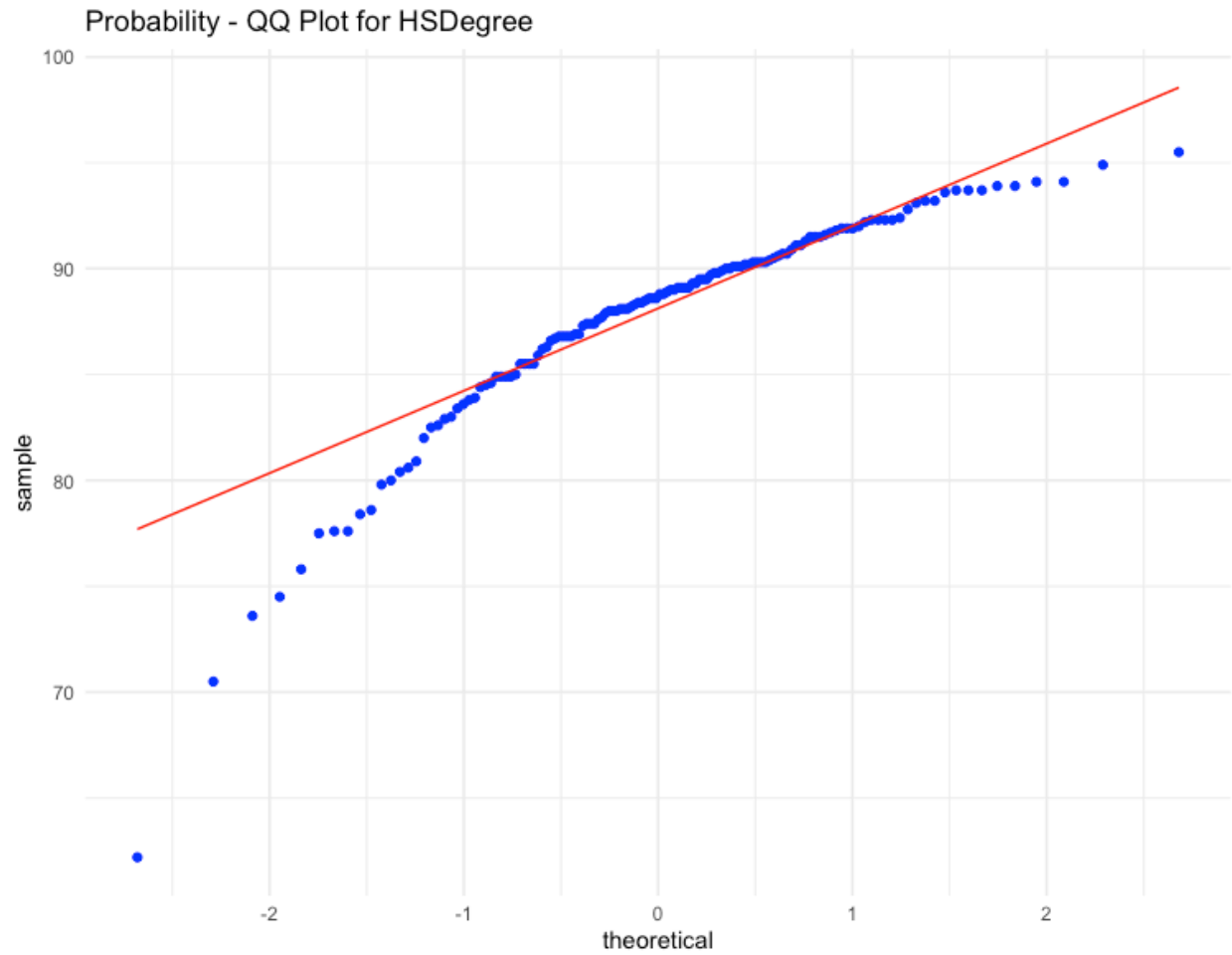
```



```
> ggplot(survey_df, aes(HSDegree)) + geom_density(aes(HSDegree),  
colour="blue") + xlim(c(60,100)) + ggtitle('County with  
HSDegree') + xlab('% Age of Population with HSDegree') +  
ylab('Density of Counties')
```



```
> ggplot(survey_df, aes(sample = HSDegree)) +  
  geom_qq(col='blue') + geom_qq_line(col='red') +  
  ggtitle("Probability - QQ Plot for HSDegree")
```



```
> library(pastecs)
> stat.desc(cbind(survey_df$HSDegree), basic = FALSE, norm =
TRUE)
```

	V1
median	8.870000e+01
mean	8.763235e+01
SE.mean	4.388598e-01
CI.mean.0.95	8.679296e-01
var	2.619332e+01
std.dev	5.117941e+00
coef.var	5.840241e-02
skewness	-1.674767e+00
skew.2SE	-4.030254e+00
kurtosis	4.352856e+00
kurt.2SE	5.273885e+00
normtest.W	8.773635e-01
normtest.p	3.193634e-09

```
>
```